Hidden Valley Higgs Search

Toward a Background Estimate

Higgs Group

Shawn Kwang
Mel Shochet

University of Chicago

Introduction

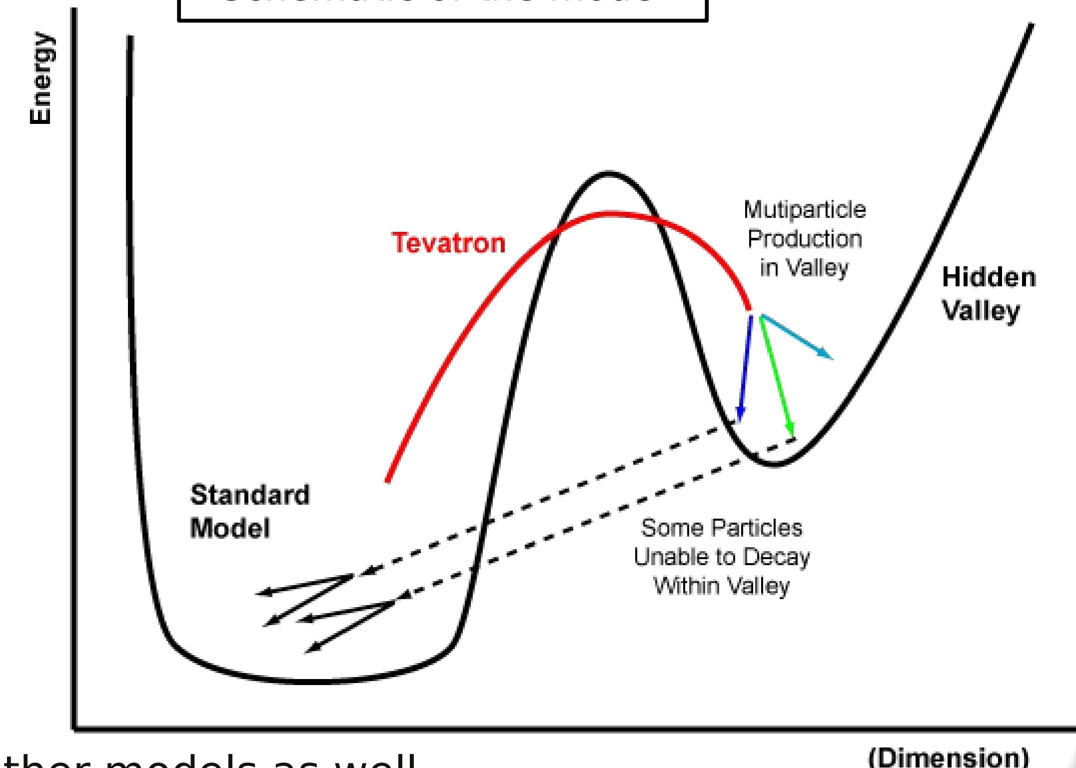
- Brief review of the Hidden Valley model.
- The big picture and algorithm flowchart.
  - Event Selection

- Determining the b-tagging probability of the jets in the ZBB trigger data.
- Determining the flavor composition of the jets in the ZBB trigger data.
  - Along the way, determine the flavor composition of JET trigger data.
- How to characterize a secondary vertex (p.d.f.s in more detail).
  - Build p.d.f.s / histograms from data.
  - Perform heavy flavor subtraction in order to isolate light-quark p.d.f.s
- Construct pseudo events from the real data.
- Verifying the pseudo events are valid.
- Using the pseudo events to make a background estimate.
  - Comparing the pseudoevents to signal MC.
  - Making basic cuts to optimize signal efficiency.
  - Preliminary background estimate.

Hidden Valley

- Energy from collisions enter into the new sector.
- It is transformed into multiple particles through the dynamics of the new sector.
  - These valley-particles (or v-particles) behave in the same way as SM particles.
  - They obey a “v-QCD,”
  - Most likely decay is a v-π.
- Some of these particles decay back into SM particles.

Schematic of the model



- This model can co-exist with other models as well.
  - SUSY, technicolor, etc.
- It may help in the search for the Higgs.
  - The Higgs may decays into long-lived neutral v-particles, which are heavy and meta-stable. They would decay at a displaced vertex.
  - These would then decay into the heaviest SM fermion available (b-quarks).
- Because this sector is dark, there may be Dark Matter/Astrophysics connections as well.
- In some models (see Kaplan, Luty, Zurek) $c\tau$ for the heavy metastable particle could be of order 1 cm.

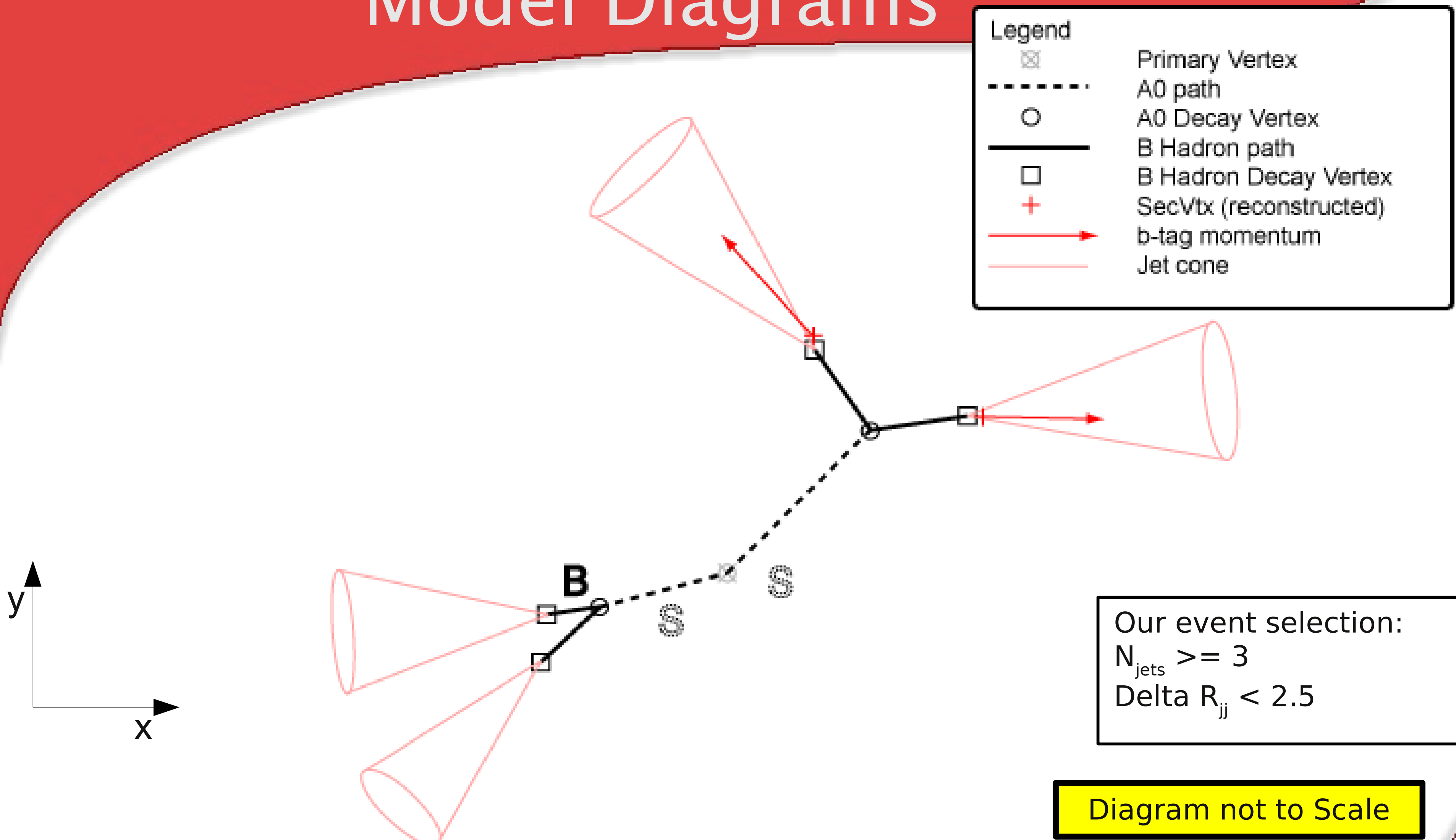
Big Picture

- The trigger in which we are searching for this signature is the ZBB trigger data.
- We are building p.d.f.s for single jets to form the background estimate.
  - We split the jets into four E_T bins, L5 corrected,
  - and the number of SVT tracks that pass the ZBB trigger requirements.
    - split into 3 SVT track bins
  - Why? The ZBB trigger is our signal trigger. It has a SVT track requirement. We separate our jets into these bins to account for the fact that this SVT requirement sculpts the distributions.
    - Details of the ZBB trigger are in the Backup slides.
- TStnSVF is a (T)Stntuple Secondary Vertex Finder.
  - The algorithm is the same as SecVtx, but the input data is from the Stntuple instead of Production data.
  - The code allows the user to change the parameters of the module in the same way as the tcl talk-tos for SecVtx.
  - Adjustments can be made to the jet, track, and vertex cuts used by the algorithm.
  - We run this b-tagger over 20 max d_0 cuts for the tracks in the jet, in order to find a d_0 cut which maximizes the efficiency for finding a signal.

Analysis Strategy

- We want to use real data to estimate our background.
  - Build p.d.f.s for background jets for a couple of variables
    - Mundane b background: QCD bb, ttbar, Z etc.
    - Mundane c background: QCD cc, Z
    - Light flavor background: QCD qq/gg
    - (Others such as tau hadronic)
  - Use data triggers when possible to build these p.d.f.s.
    - Muon/Electron calibration data, which is rich is heavy flavor jets
    - Pythia QCD cc MC
    - Single Tower 5, jet data, for light-quark and gluon jets
  - These p.d.f. are per jet (not per event).
  - These per jet p.d.f.s can be applied to multijet QCD production, either data or MC, to estimate the final background and decide on cuts.
- What p.d.f. variables?
  - The variables are those dealing with the secondary vertex. Specifically characterizing the secondary vertex's position and momentum.
  - More details are given later.

Model Diagrams



Our event selection:
$N_{jets} >= 3$
Delta $R_{jj} < 2.5$

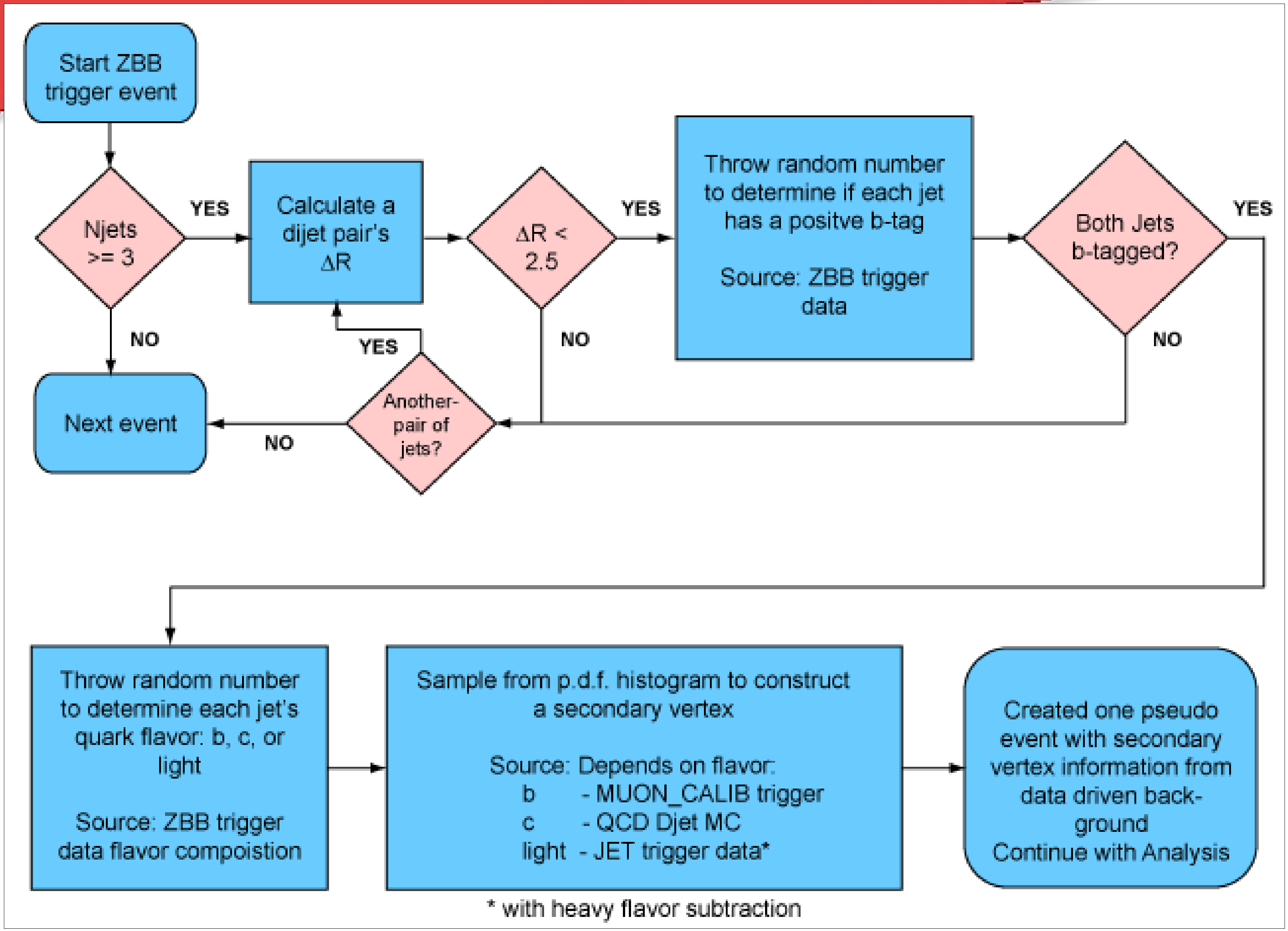
Diagram not to Scale

Here the Higgs decays at the primary vertex (the X). S represents the heavy pseudoscalar with a long lifetime, which decays into bbbar pairs.
The pink cones represent the hadronization of the B hadrons into jets.
The red represents reconstructed secondary vertices and their corresponding momenta.
Black is the “truth” information.

Signal Event Selection

- We will be looking for events with central b-tagged jets, with a relatively low E_T requirement. "Signal Region"
  - All jets are required to have:
    - $E_T > 20$ GeV, corrected at Level-5
    - $|\eta|<1.0$
  - Jet multiplicity: $N_{jet} \geq 3$
  - For the dijet system, require that it be in a region that would be populated by signal.
    - $\Delta R < 2.5$
- A "Control Region" is defined which contains events orthogonal to the Signal Region,
  - Two tight central jets ($N_{jet} = 2$)
  - A third jet with Level 5 corrected $E_T < 15$ GeV.

Flowchart



This flowchart shows the background estimate / pseudo event generating algorithm.
Each step in the blue rectangles will be discussed on the following slides.

B–tag probability

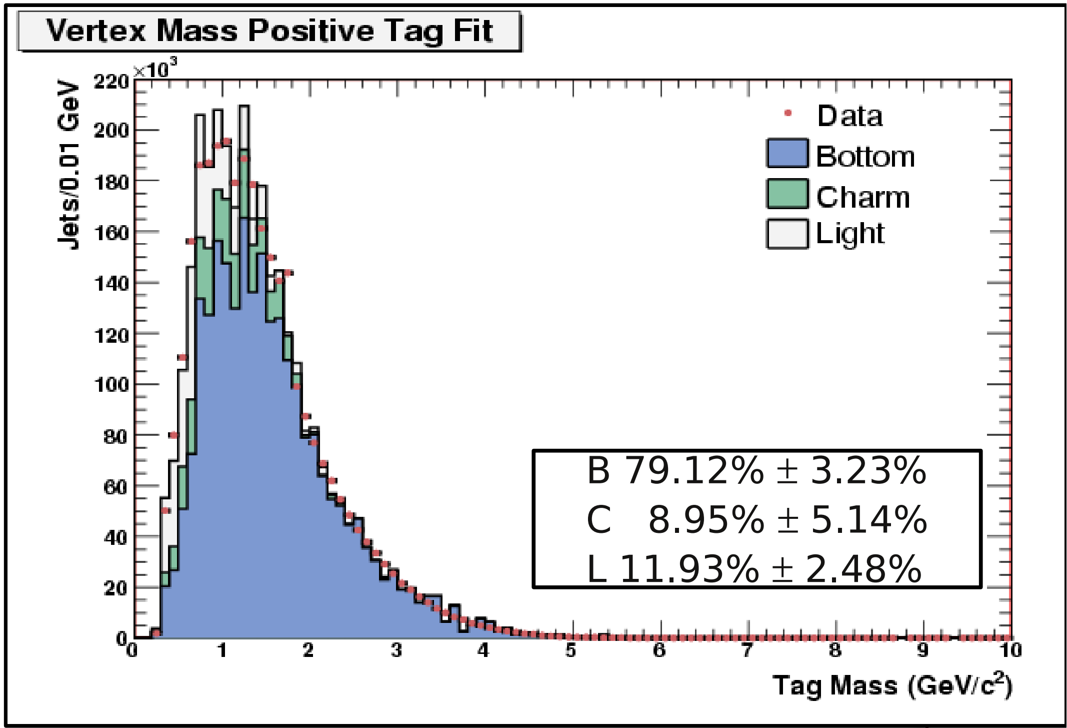
Determining the b-tagging probability of the jets in the ZBB trigger data

- Fairly simple, count the number of pre-tagged jets in given Et and #SVT track bin.
- Count the number of b-tagged jets in the same bins.
  - Done for each of the twenty max d_0 cuts set for TStnSVF.
- Divide the number tagged jets by the number pre-tagged jets and you get a probability for each bin.
- Event selection: ZBB trigger data: dataset ezbbxx
  - Inclusive jet selection
  - Data is from p0-25.

Flavor Composition

Determining the flavor composition of the jets in the ZBB trigger data

- We want to know the flavor composition of the ZBB trigger data.
- For each positive b-tagged jet plot the vertex mass of the secondary vertex.
  - In each Et and #SVT track bin, for all twenty max d_0 cuts
- For templates we turn to the QCD dijet MC.
  - Templates were formed of positive b-tagged jets in a similar fashion, split into b-quark, c-quark, and light flavor jets.
- We use ROOT's TFractionFitter to fit the MC templates to the data to give us the fraction of b, c, and light flavor quarks in each bin.



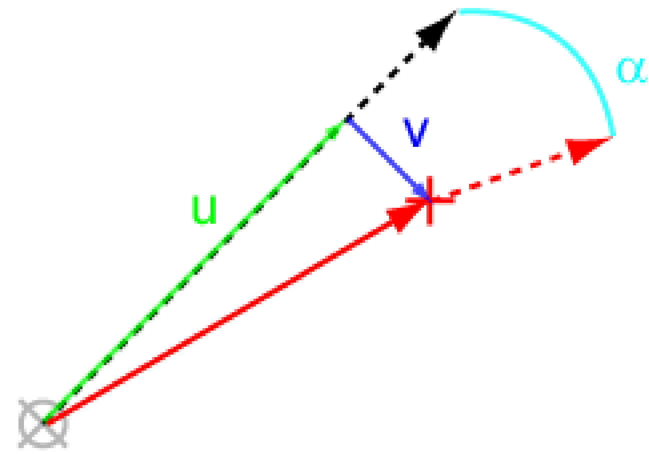
ET bin 30-70 GeV, 1 SVT track, max d_0< 0.30 cm

- We also perform the flavor composition on the JET trigger data (JET_20, JET_50, etc.)
  - This is later used in to subtract heavy flavor from the JET trigger data in order to obtain a light flavor sample of positively tagged jets.

Building p.d.f.s

How to characterize a secondary vertex (p.d.f.s in more detail)

- There are three p.d.f. variables
  - u – the L_{xy} component vector parallel to the jet axis
  - v – the L_{xy} component vector perpendicular to the jet axis
  - α – the angle between the secondary vertex momentum and the jet momentum
- We are working in 2 dimensions.
- These three variables encapsulate the information about a single secondary vertex.
- These are plotted in a three dimensional histogram, this preserves all correlations between these variables.



Black Dashed Line	Jet Momentum
Red Cross	Secondary Vertex
Red Solid Line	Sec Vertex L_{xy}
Red Dashed Line	Sec Vertex Momentum
Green Line (u)	Parallel component of L_{xy}
Blue Line (v)	Perp. component of L_{xy}
Cyan Arc (α)	Angle b/w the two momenta

Data Sources

Build p.d.f.s / histograms from data (when possible)

- B quark p.d.f.s
  - We use the muon calibration trigger path (MUON_CALIB8) and the strategy from the B-tagging scale factor (BTSF) analysis.
  - Heavy flavor jets identified in jet pairs, where one jet has a muon (lepton-jet) and is tagged, and the second is also tagged (away-jet).
  - In order not to bias ourselves with soft lepton b-jets, we take the away-jet as the source for our p.d.f.s.
    - Data is from bmclxx, p0-25.
- C quark p.d.f.s
  - Since there is no dedicated charm trigger, we use the QCD dijet MC
    - Pythia QCD dijet: bt0s[qrst]b, bq0s[qrst]c, and bq0s[sn]d
    - E_T cuts (L5 corrected) are employed to remove any turn on bias from the p_T^{hat} cut.
    - These datasets are normalized by their L5 corrected jet E_T to produce a smoothly falling spectrum.
  - The QCD dijet MC is also used to calculate the flavor composition.
- Light flavor p.d.f.s
  - We use the jet trigger data because of its large statistics.
    - SINGLE_TOWER5, JET_20, JET_50, JET_70, JET_100
    - Data is from p0-25.
    - E_T cuts (L5 corrected) are employed to remove any turn on bias.
    - These datasets are also normalized by their L5 corrected jet E_T to produce a smoothly falling spectrum.

Heavy Flavor Subtraction

Perform heavy flavor subtraction in order to isolate light-quark p.d.f.s

- Earlier I mentioned that we calculated the flavor composition of the JET triggers.
- For positively tagged jets in the JET trigger, many of the jets are b and c quarks.
- We remove these by subtracting off the heavy flavor contributions.
- The source of the b and c shapes are the QCD dijet MC samples.

Outline

- Brief review of the Hidden Valley model.
- The big picture and algorithm flowchart.
  - Event Selection
- Determining the b-tagging probability of the jets in the ZBB trigger data.
- Determining the flavor composition of the jets in the ZBB trigger data.
  - Along the way, determine the flavor composition of JET trigger data.
- How to characterize a secondary vertex (p.d.f.s in more detail).
  - Build p.d.f.s / histograms from data.
  - Perform heavy flavor subtraction in order to isolate light-quark p.d.f.s
- Construct pseudo events from the real data.
- Verifying the pseudo events are valid.
- Using the pseudo events to make a background estimate.
  - Comparing the pseudoevents to signal MC.
  - Making basic cuts to optimize signal efficiency.
  - Preliminary background estimate.

Using the algorithm shown earlier and the information shown in the previous slides, we now construct the pseudo events that form our background estimate.

Diagrams ψ

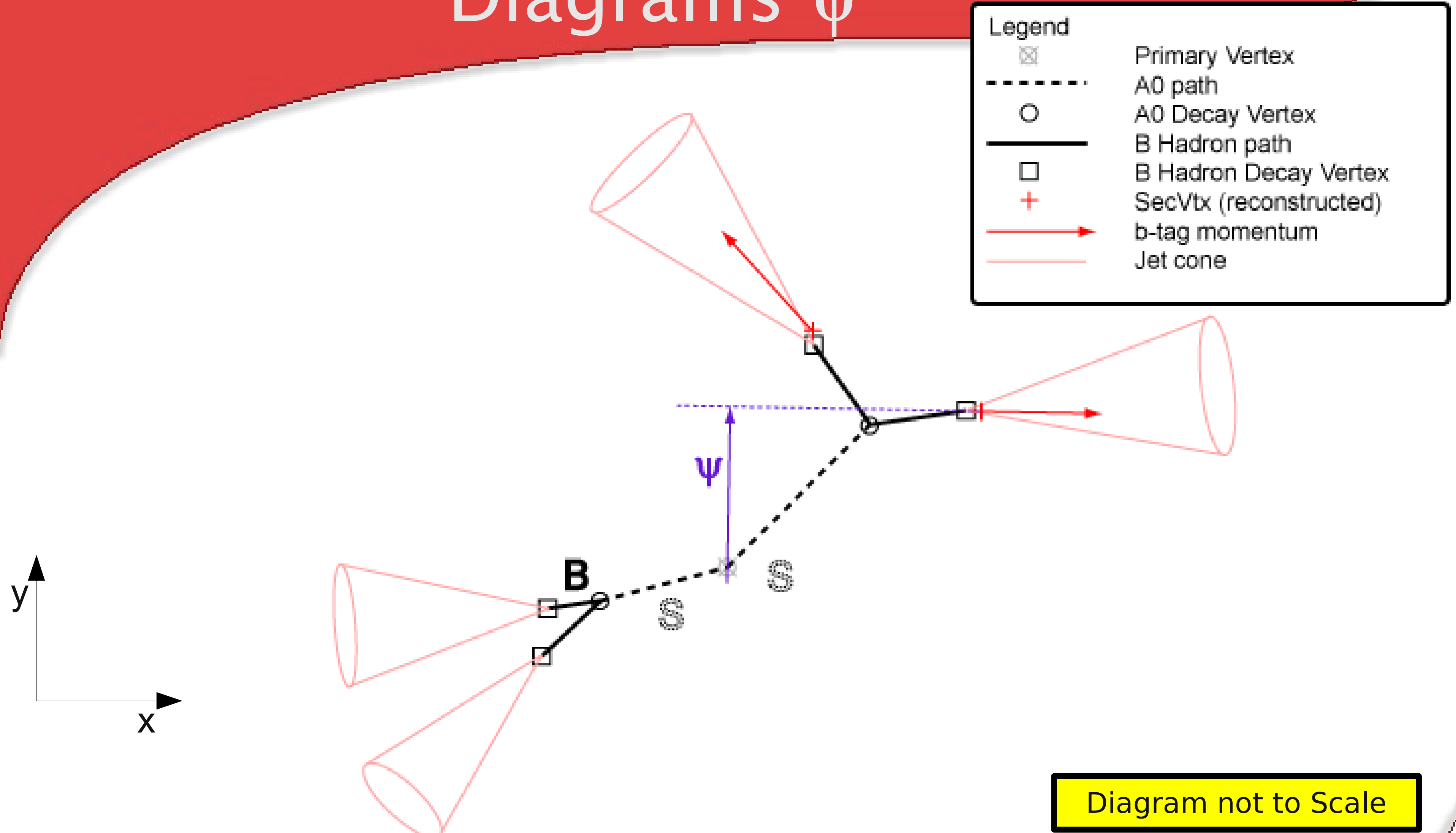
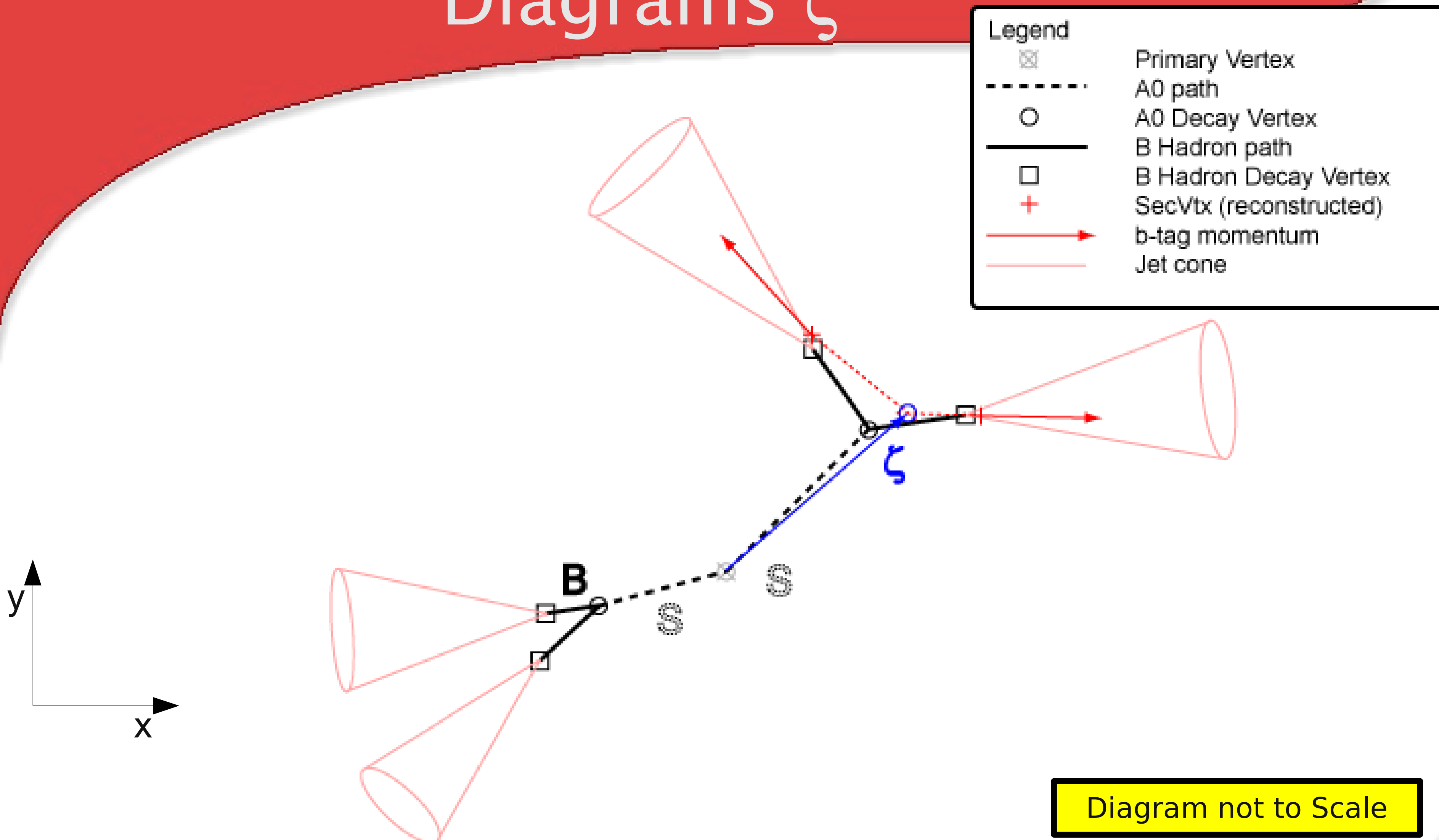


Diagram not to Scale

ψ is the impact parameter of a jet with a secondary vertex.

This is in two-dimensional space.

Diagrams ζ



ζ is the reconstructed decay distance of the heavy pseudoscalar S (A_0). It requires two tagged jets.

This is in two-dimensional space.

Background Diagrams

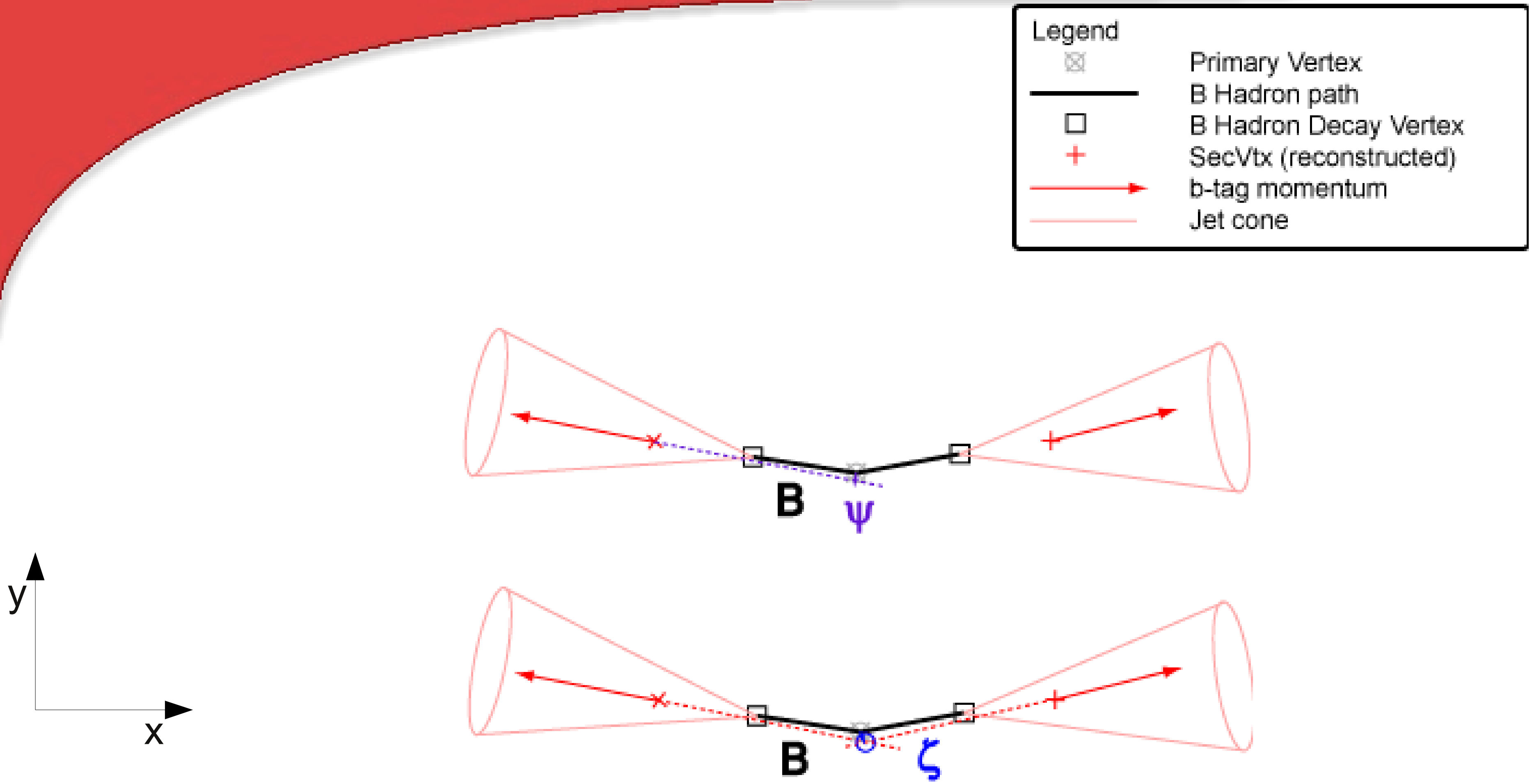
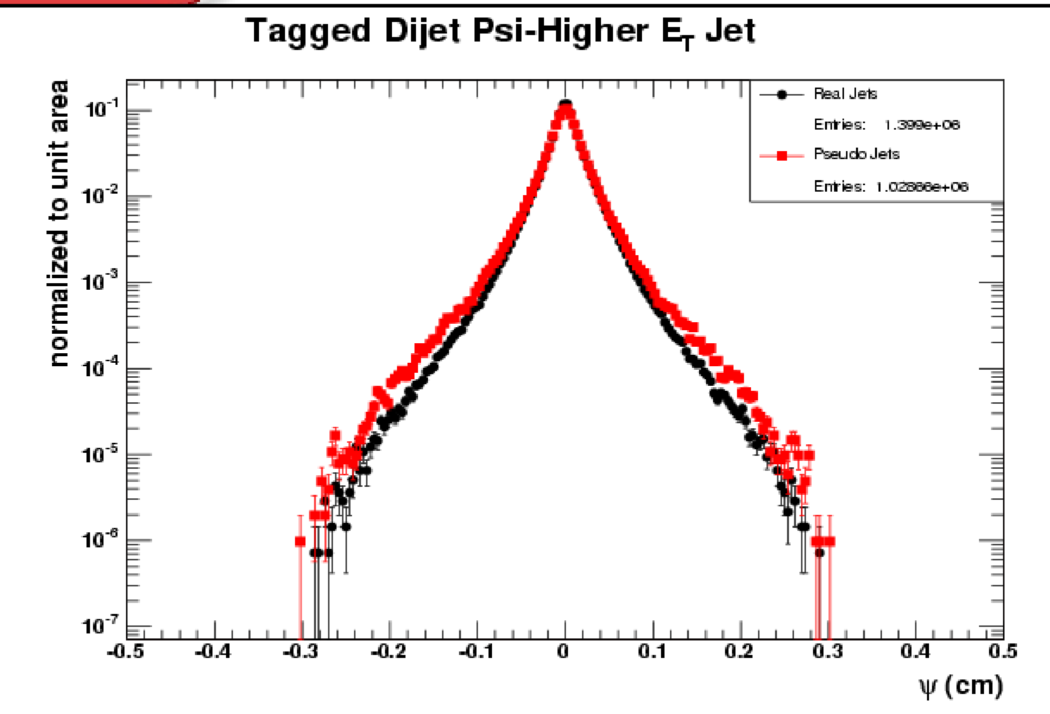


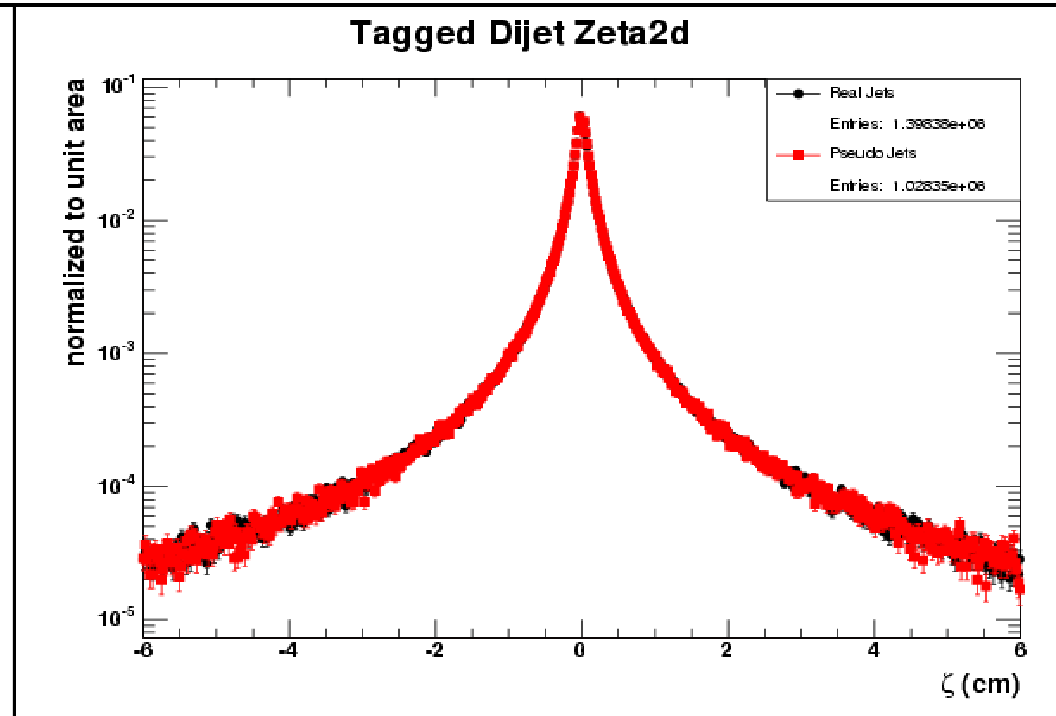
Diagram not to Scale

Here is a typical QCD di-jet event with two b quarks (b & bbar) decaying into two B hadrons. Each has a reconstructed secondary vertex represented by a red cross. Both ψ/ζ are very small for these background events.

Verifying the Algorithm

Verifying the pseudo events are valid





- To verify this algorithm we use a 2-jet Control Region
  - The event selection is 2 jets, with a third jet with (L5) E_T < 15 GeV.
  - The red points are from the pseudo jets, while the black points are the real jets in the event.
  - What is plotted are all events from all bins for a particular max d_0 cut

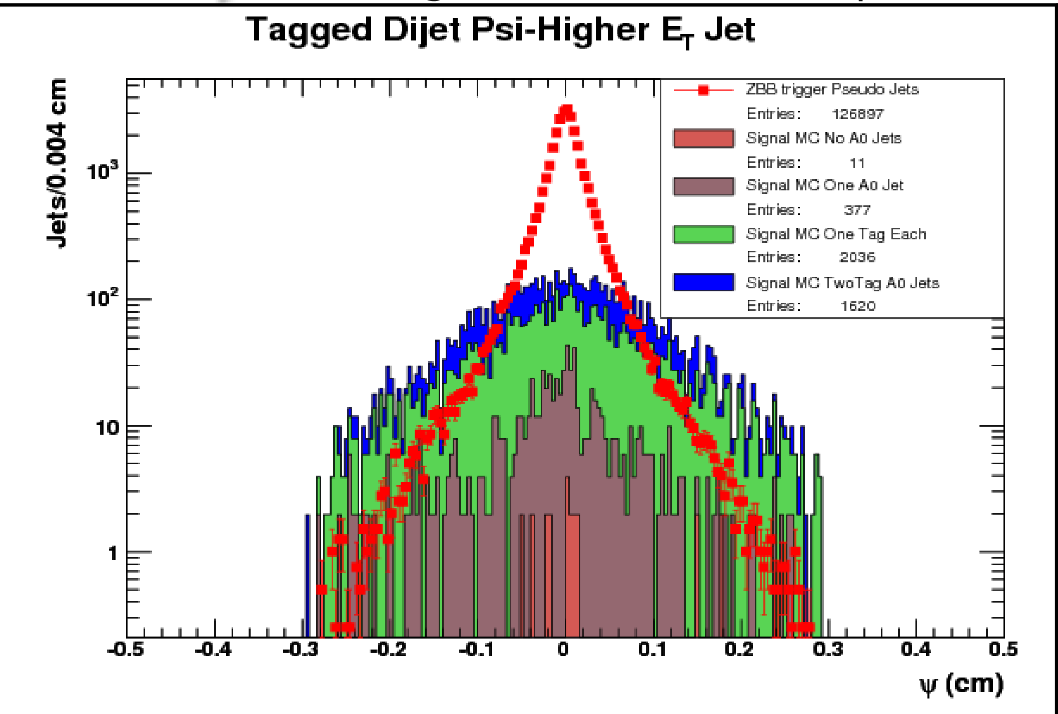
($|d_0| < 0.30$ cm).

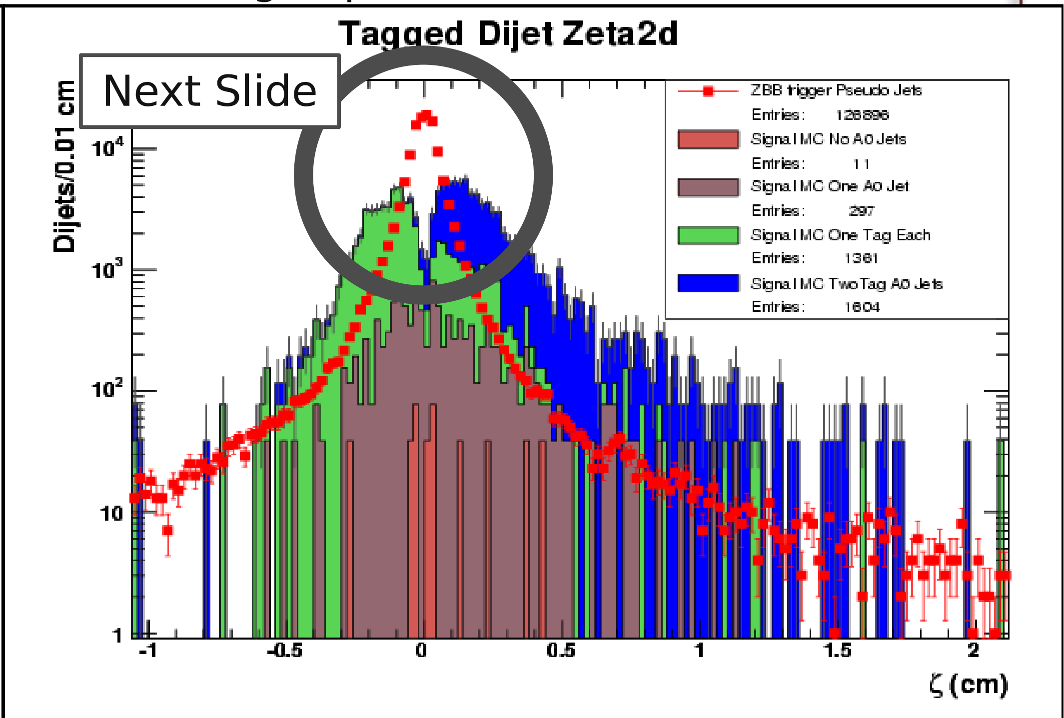
- The discrepancies in the shapes along the tail (of ψ) will result in a systematic uncertainty.
  - It will also result in an overestimate of our background.

Comparing Signal to Background

Comparing the pseudoevents to signal MC

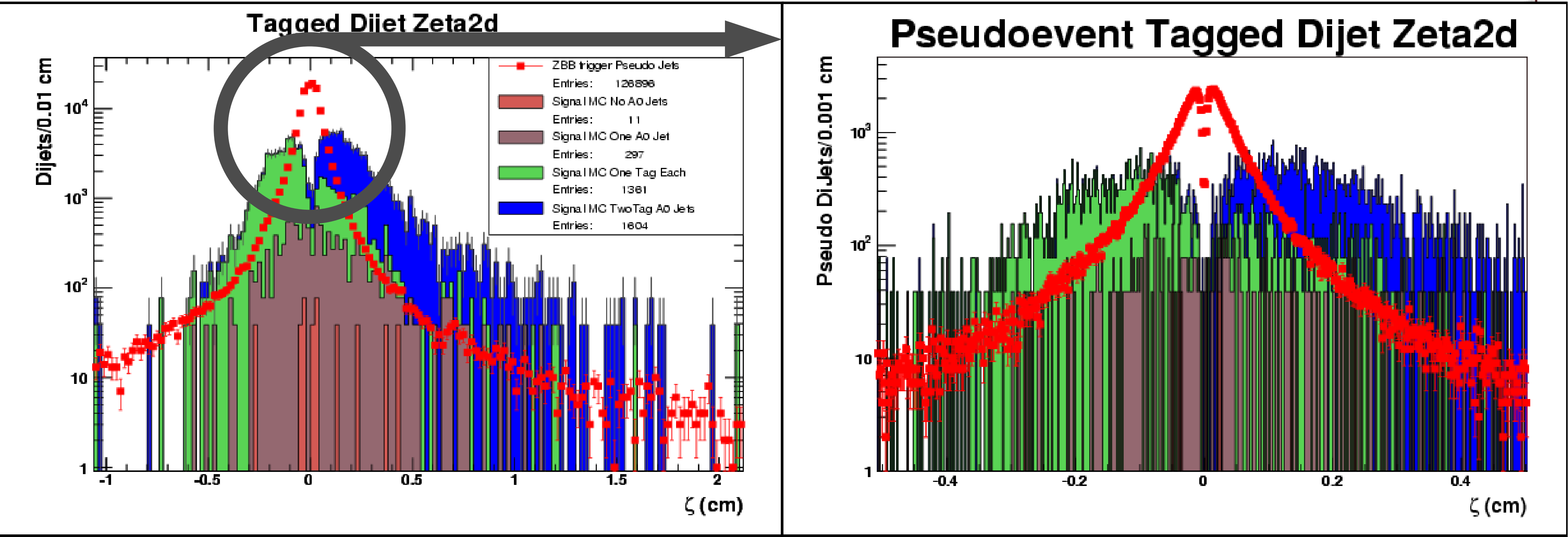
- Signal MC was generated using Pythia, with at number of different mass points for the Higgs and the heavy pseudo-scaler.
  - $H_0 \rightarrow A_0A_0 \rightarrow$ bb, bb: the A_0 represents the long-lived pseudoscalar.
  - What's shown here is H_0=130 GeV and A_0=40 GeV with lifetime 1.0 cm.
  - The max d_0 cut is $|d_0| < 0.30$ cm.
- Compare the signal MC to the pseudo events generated from the ZBB trigger data – background estimate data, 3 jet Signal region.
  - The signal MC area has been normalized to the pseudo data.
- Looking to make some simple cuts for a counting experiment.





Signal vs Background

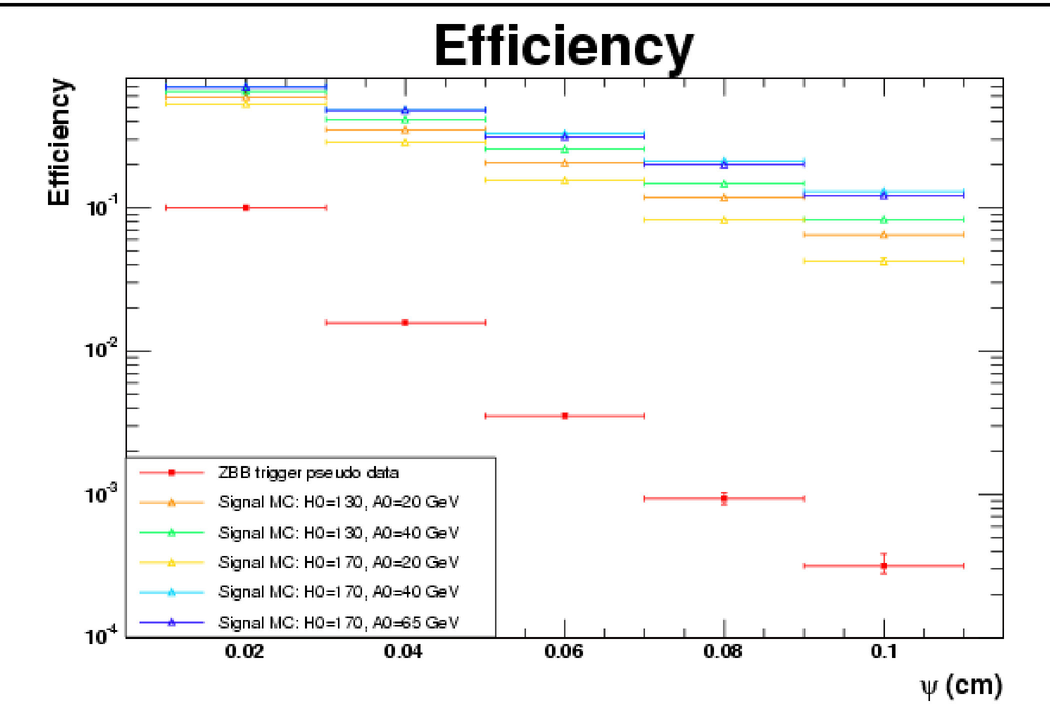
- ζ is a 2 dimensional distance: from the primary vertex, to the intersection of the two b-tag directions.
- Left: same plot as last slide
- Right: zoomed in, with 10x more bins.
  - Phase space
  - Resolution effect
- There is more phase space for ζ to exist away from the primary vertex, and the signal MC by construction has larger ζ, the result is the "hole" you see below.

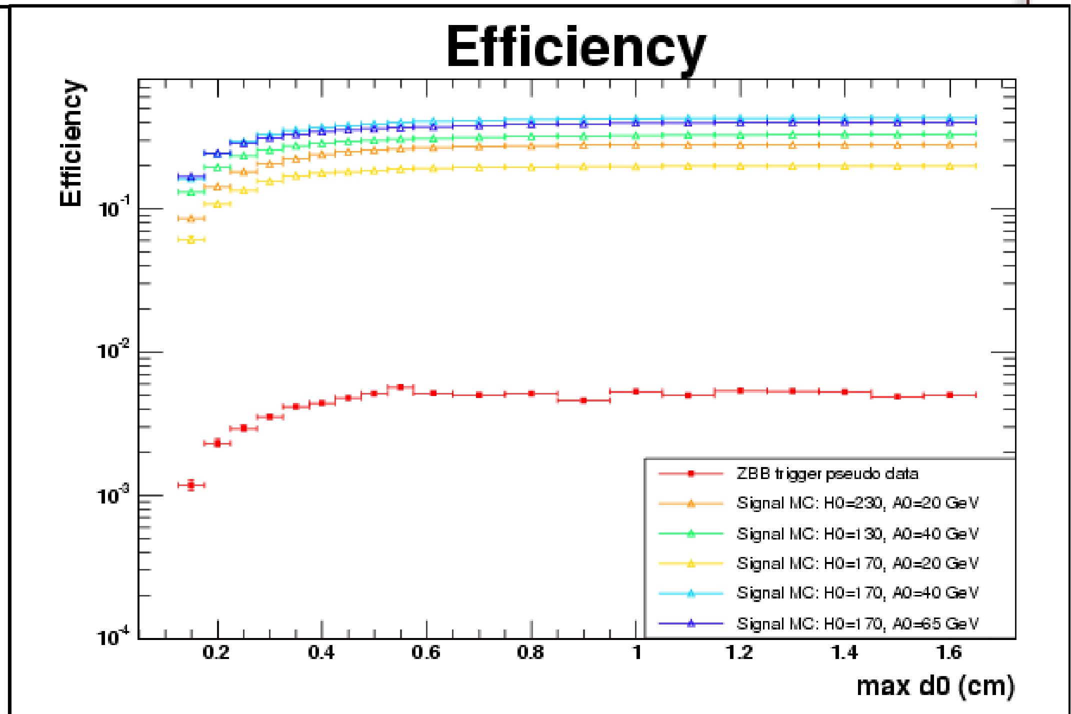


Signal Efficiency

Making basic cuts to optimize signal efficiency

- Make a cut on ψ, see how it affects the signal efficiency.
- Left: Efficiency on the background estimate and signal MC (five different H_0 and A_0 masses) as a function of ψ.
  - For a max d_0 of $|d_0| < 0.30$ cm.
- Right: Efficiency for a single cut ($|\psi| > 0.06$ cm) for all max d_0 cuts.
  - This is a cut on both b-tagged jets.

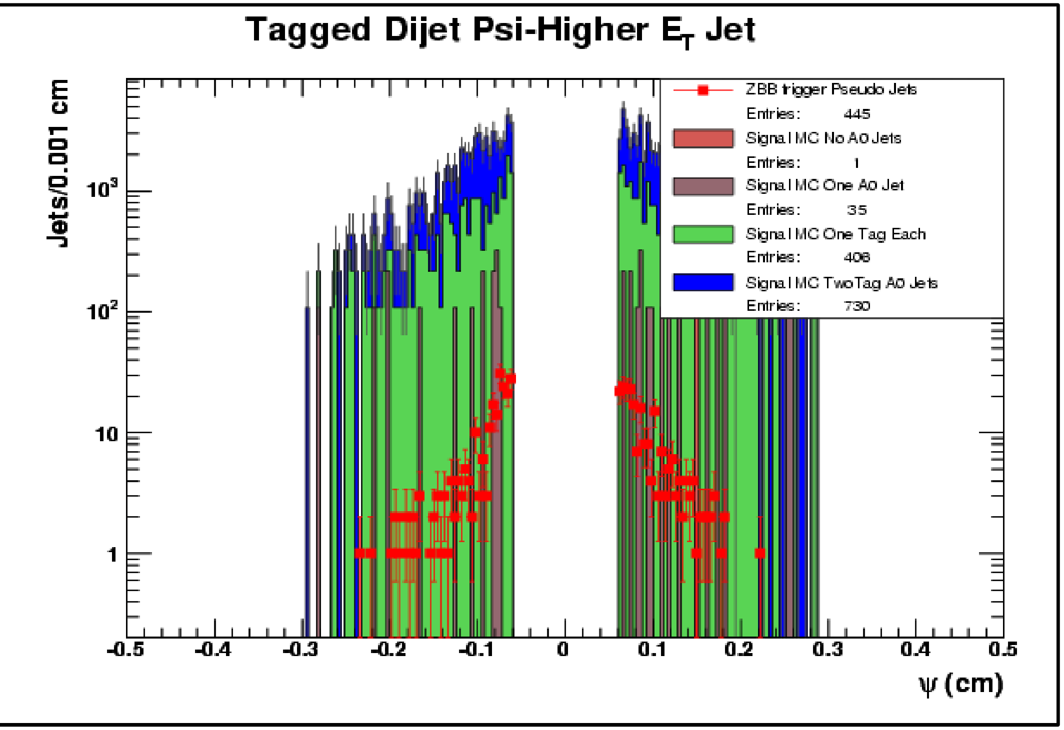




Preliminary Estimate

Preliminary background estimate

- This background estimate is a proof of concept. Our actual results may change.
- A simple cut on |ψ|, demanding both b-tagged jets have |ψ| > 0.06 cm results in a reduction from 126897 to 445 background events.
  - The max d_0 cut is $|d_0| < 0.30$ cm.
  - Below, there is no cut on ζ.
  - Signal MC is normalized in the same way as before, showing the result of the cut.



Future Tasks

- We need to determine how robust the background estimate is, i.e. what are its systematic uncertainties.
  - Flavor composition uncertainty propagates as a systematic.
    - There is a different MC track efficiency which affects the shapes of the vertex mass templates. These may affect the background estimate by propagating through the flavor composition uncertainty.
  - Three dimensional histogram binning.
    - The p.d.f.s for the background estimate are binned data. A different coarser binning may affect our background estimate.
  - The differences shown in the Control region between the pseudo jets and the real jets.
    - These differences represent an uncertainty in how we have modeled the background.
- Select our final cuts based on retaining as much signal efficiency as possible.

Conclusion

- Shown how to construct pseudo events from real data. Thus constructing data driven background estimate.
  - This background estimate is automatically normalized to the luminosity of the real data it was generated from.
- Compared the background estimate to our signal MC to make some basic cuts.
  - Trying to preserve some signal efficiency while eliminating the background.
  - Looking for an excess of events.
- Finally, made some preliminary cuts and showed how it affects the background estimate/pseudo data.

Backup Slides

- ZBB Trigger Details
- Jet binning
- Additional Diagrams

ZBB Trigger

- Details of the trigger in the trigger table:
  - L1 :
    - one central tower with $E_T > 5$ GeV
    - two XFT tracks, $p_T^1 > 5.48$ GeV, $p_T^2 > 2.46$ GeV
  - L2 :
    - veto events w/ clusters with $E_T > 5$ GeV, $|\eta| > 1.1$
    - requires two clusters $E_T > 5$ GeV, $|\eta|<1.1$ which have $9 < \Delta\text{Wedge} < 12$
    - two SVT tracks with $p_T > 2$ GeV, $d_0 > 160$ microns, $d_0 < 1000$ microns, $\chi^2 < 12$,
      - $150 < \Delta\phi < 180$ "Opposite Side"
      - $0 < \Delta\phi < 30$ "Same Side"
      - This triggers on displaced tracks in the event.
  - L3:
    - two R=0.7 jets with $E_T > 10$ GeV, $|\eta| < 1.1$
    - two SVT tracks with $p_T > 2$ GeV, $d_0 > 160$ microns, $d_0 < 1000$ microns, $|\eta| < 1.2$
    - two tracks with $p_T > 1.5$ GeV, $d_0 > 130$ microns, $d_0 < 1000$ microns, $|\eta| < 1.2$, IP significance $Sd_0>3$, $\Delta z < 5$ cm
  - Dynamically Prescaled Trigger
  - This is for the latest trigger "chunk," #17. Chunks 10-16 are nearly the same, with minor changes in the cut values, but the structure is the same.

Jet Binning

- Out jets are split into 12 different bins.
  - Four bins of E_T: [20,30) [30, 70) [70,110) [110,200)
  - Three bins of # SVT tracks: 0, 1, (>=2)
- The E_T bins are the result of the jet trigger
  - SINGLETOWER5 → [20,30)
  - JET_20 → [30,70), etc.
- The # SVT track bins are split as such because the ZBB trigger requires two SVT tracks in the event, not per jet.
  - In order to account for the differences in tag probability, flavor, and b-tag kinematics, it is necessary to split our QCD jet sample into the same SVT tracks requirements that the ZBB trigger uses.

Diagrams ζ

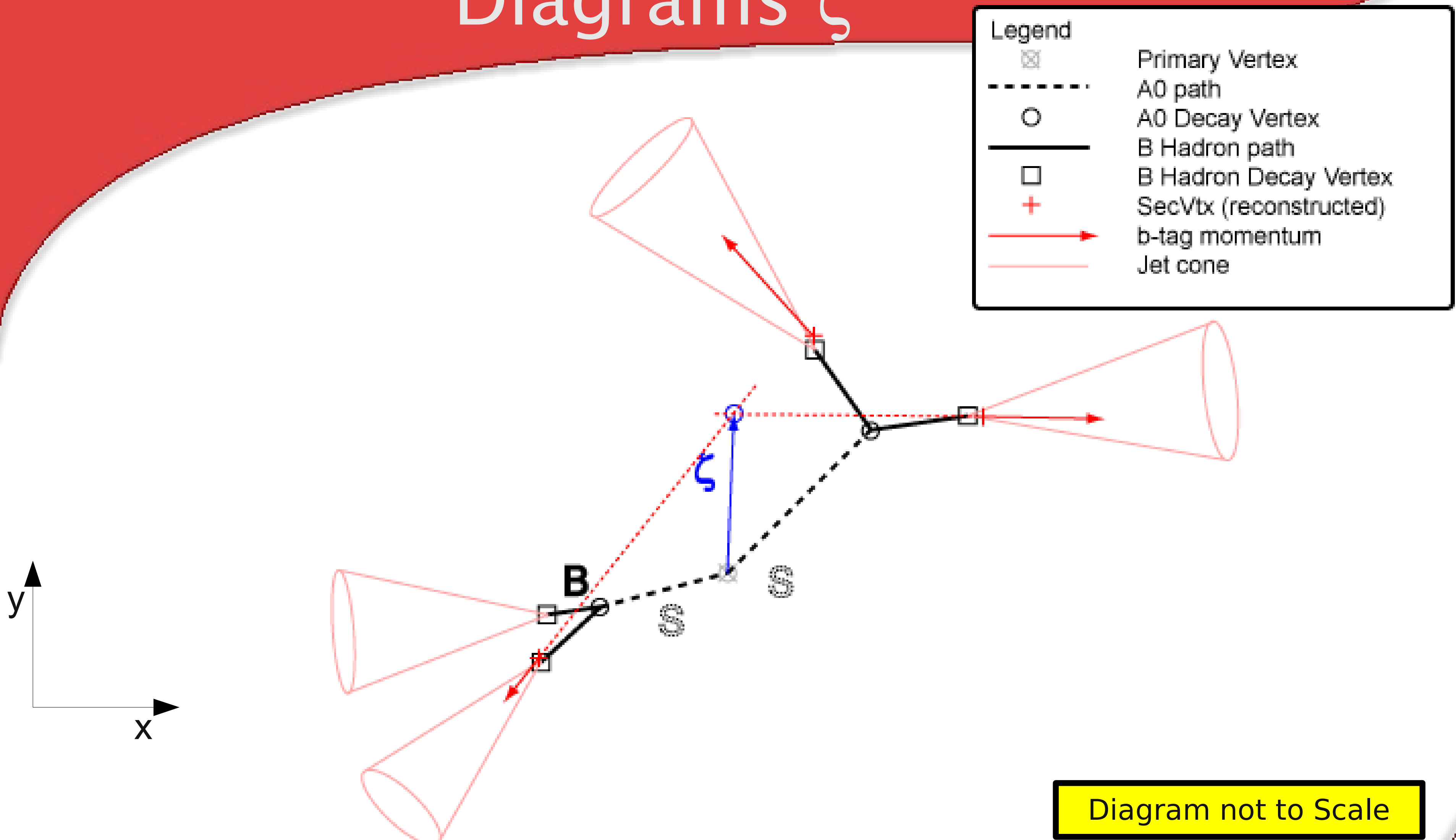


Diagram not to Scale

ζ can be calculated for two b-tagged jets from different A_0 decays. This is called "OneTagEach".

This is in two-dimensional space.

Diagrams ΔS

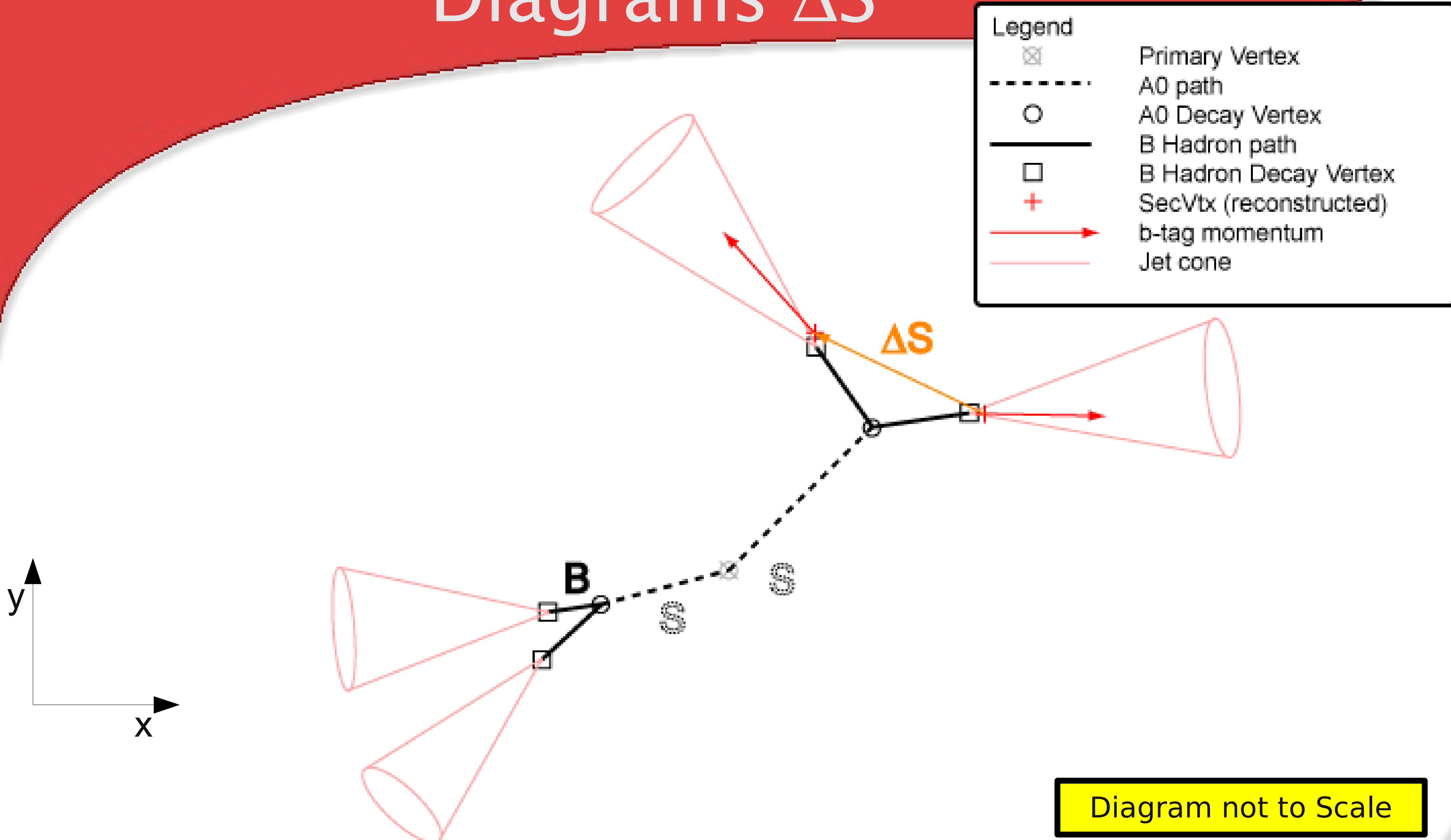


Diagram not to Scale

ΔS is the two-dimensional distance from one secondary vertex to the other.
It should be rather small for our signal.

This is in two-dimensional space.

Zeta after cuts

- ζ after the |ψ| > 0.06 cm cut discussed earlier.
  - Hole is caused by the |ψ| cut
  - The signal MC shows a significant excess in positive ζ.